Molecular evolution of prolactin in Chiroptera: accelerated evolution and a large insertion in vespertilionid bats

Molecular evolution of prolactin in Chiroptera: accelerated evolution and a large insertion in vespertilionid bats

Michael Wallis

Biochemistry and Biomedicine Group, School of Life Sciences, University of Sussex, Brighton, BN1 9QG. UK

Corresponding author: address as above; email: m.wallis@sussex.ac.uk

Short title: Molecular evolution of prolactin in bats

Keywords: prolactin, receptor, Chiroptera, inserted sequence, molecular evolution, evolutionary rates

Declaration of Interest: None

Abstract

Pituitary prolactin (PRL) shows an episodic pattern of evolution in mammals, with a slow underlying rate (near stasis) and periods of rapid change in some groups. PRL evolution in bats, the second most speciose mammalian order, has not previously been studied, and is examined here. Slow basal evolution of PRL is seen in some bats, particularly megabats, but in most microbat groups evolution of PRL is more rapid. Accelerated evolution of PRL is particularly notable in the family Vespertilionidae, where analysis of nonsynonymous and synonymous substitutions indicates that it reflects adaptive evolution/positive selection. Remarkably, vespertilionid bats also show a large sequence insertion, of variable length, into exon 4 of PRL, giving a protein sequence 18-60 amino acids longer than normal, with the longest insertions in bats of the genus *Myotis*. An equivalent insertion has not been reported in PRL of any other vertebrate group. In the 3-dimensional structure of the complex between PRL and the extracellular domain (ecd) of its receptor (PRL:$PRLR_2$) the inserted sequence is seen to be introduced in the short loop between helices 2 and 3 of PRL; it is far removed from the receptor-binding sites, and may not interfere with binding. The ecd of the receptor also shows variable rates of evolution, with a higher rate in the Vespertilionidae, but this is much less marked than for the hormone. The distribution of substitutions introduced into PRL during vespertilionid evolution appears to be non-random, and this and the evidence for positive selection suggests that the rapid evolution and insert sequence introduction were associated with a significant change in the biological properties of the hormone.

1. Introduction

Pituitary prolactin (PRL) is a protein hormone that plays an important role in regulating lactation in mammals, and a range of other physiological processes both in mammals and non-mammalian vertebrates (Clarke and Bern, 1980; Horseman and Gregerson, 2014; Shillingford and Hennighausen, 2001). PRL is structurally similar to pituitary growth hormone (GH), and is a member of the 4-helical bundle cytokine superfamily (Teilum et al., 2005). Like other members of this family PRL acts by binding to a membrane-associated receptor with a single transmembrane helical domain. The PRL receptor (PRLR) exists at least partly as a preformed homodimer, though binding of PRL may facilitate dimerization (Clevinger et al., 2009). Following ligand binding, conformational changes of the receptor lead to activation of associated signalling molecules, particularly JAK2 kinase (Clevinger et al., 2009; Horseman and Gregerson, 2014). The structure of the extracellular domain (ecd) of the receptor dimer bound to PRL (PRL:$PRLR_2$) has been determined by crystallography (Broutin et al., 2010; van Agthoven et al., 2010). The conformation of the intracellular domain (icd) is less well understood, though recent studies suggest that it may be largely unstructured, providing a scaffold for sequence motifs that can bind specific signalling molecules (Bugge et al., 2016; Haxholm et al., 2015).

In mammals, both PRL and GH show an interesting pattern of molecular evolution in which a generally very slow basal evolution (near stasis) is punctuated by episodes of rapid change (Forsyth and Wallis 2002; Wallis 1981; 1994; 2000; 2009;Wallis et al., 2005). For GH, one such episode occurs within the Chiroptera (bats) (Wallis 2008), but evolution of bat PRL has not been reported. The Chiroptera is a large mammalian order, the second most speciose after Rodentia. It shows a number of remarkable specializations, including flight, and in some species echolocation and hematophagy (feeding on blood). Some bat species have an exceptionally long life-span in relation to size (Dammann, 2017)

Recent developments in comparative genomics have led to the availability of complete or partial genome sequences for over 100 mammals, including many bats. Transcriptomic data are also available for many bat species. Such data make possible the characterization of sequences of PRL and its receptor from about 20 bat species, allowing a detailed evaluation of the molecular evolution of the

PRL/PRLR pair in this mammalian order. This showed that PRL in Chiroptera shows variable evolutionary rates, as in some other mammalian groups, and also, in vespertilionid bats, introduction of a long insertion, of a kind unseen in any other vertebrate group.

2. Methods

2.1. Sequences

The sources of the sequence data used in this study are given in Supplementary Table 1. For a number of bat species genomic assemblies are available at www.ncbi.nim.nih.gov, and sequences for PRL and PRLR genes were identified using BLAST searches (Altschul et al., 1990) on the appropriate whole-genome shotgun contig database. In other cases genomic assemblies were not available, but PRL and/or PRLR sequences could be obtained from SRA genomic or transcriptomic databases (https://trace.ncbi.nlm.nih.gov/Traces/sra/) using BLAST . In these cases, the appropriate database/experiment was subjected to a Blast analysis on the SRA website, with a PRL CDS sequence (initially from a related species) as Query. The algorithm chosen (megablast, discontiguous megablast or blastn) was according to the extent of difference between template and target. Other parameters were the defaults on the sra Blast website, except that the 'number of aligned sequences to display' was set to maximum. Formatting options were set to 'flat query-anchored with dots for identities' and line length 150. This retrieved sequences that were similar to the Query sequence, aligned/assembled on the template (Query), allowing derivation of the full or partial sequence of the target. Gaps were filled by 'walking up' from adjacent sequence. In all cases sequences were confirmed by searching original databases using BLAST. Coding sequences (CDS) were extracted, based on the 5-exon organization of the human PRL gene (Truong et al., 1984) and the rabbit PRLR mRNA (Edery et al., 1989), which includes 8 coding exons. Examination of bat transcriptomes showed that these are the main forms of bat PRL and PRLR mRNA.

2.2 Alignments

CDS sequences for bat PRL and PRLR were assembled using the Mesquite software, version 3.5 (Maddison and Maddison 2018), and translated to give protein sequences. Sequence alignment was carried out manually. Sequences of rabbit, dog and in some cases human PRL and PRLR were included

in alignments as outgroups. Rabbit and dog PRL and PRLR were chosen because they are strongly conserved, showing slow evolution and providing a good baseline for assessment of accelerated evolution in other groups. Human PRL and PRLR were included in the sequence alignments shown, because of the extensive structural and other work carried out on them, but were not used in the codeml analyses. Signal peptide position was assigned by comparison with PRL and PRLR precursor sequences from other species (including human).

2.3 Phylogenies

Bat phylogenies were based on Agnarsson et al. (2011) and Amador et al. (2018). Phylogenies obtained using the PRL and PRLR sequence data generally agreed well with these.

2.4. Sequence analysis - evolutionary rates

To analyse evolutionary rates for bat PRL and PRLR CDS sequences, the codeml programme in the paml package (Phylogenetic Analysis by Maximum Likelihood; Yang, 2007) was used to determine the ratio (dN/dS) of nonsynonymous substitutions (altering amino acid sequence) to synonymous substitutions (silent). For most proteins dN/dS is low, as purifying selection maintains functional sequence. For a sequence without specific function dN/dS approaches 1.0, the neutral rate of evolution. A dN/dS value significantly greater than 1.0 indicates that the sequence is undergoing rapid adaptive evolution due to positive selection, but a value lower than 1.0 does not necessarily rule out adaptive evolution (Nei and Gojobori, 1986).

CDS sequence alignments corresponding to all or subregions of the PRL and PRLR mRNAs were analysed using the codeml method (Yang, 2007), with a defined phylogenetic tree based on previously-described bat phylogenies (see Section 2.3). Variation in evolutionary rates for different branches in the phylogeny was investigated using the branch models in codeml (Yang, 1998; Yang and Nielsen, 1998). Positive selection at specific sites across branches was studied using the branch site models (Yang et al., 2005, Zhang et al., 2005). The likelihood ratio test (LRT) was used to test significance of differences between dN/dS ratios (Yang, 2007).

2.5. 3-dimensional models

Homology models for bat PRL:$PRLR_2$ complexes were constructed using the SWISS-MODEL server at https://swissmodel.expasy.org (Waterhouse et al., 2018) with the structure of human PRL bound to two molecules of rat PRLR (3npz; van Agthoven et al., 2010) as template. Models were visualized and marked up using PyMol (The PyMOL Molecular Graphics System, Version 1.8.7, Schrödinger, LLC).

2.6. *Transcription*

Expression of bat PRL and PRLR in a variety of tissues was assessed by examining transcriptomes available in the SRA database, using BLAST.

3. Results and Discussion

3.1. Bat prolactin sequences

Complete PRL sequences (CDS and protein) were obtained for 21 bat species, representing 8 bat families, and both megabats and microbats (Supplementary Table 1). In many species evidence was seen in the sequence data for modest allelic polymorphism, but in no case was there evidence for gene duplication or an extended family of PRL-like genes as seen in rodents and ruminants (Carter, 2012; Forsyth and Wallis, 2002).

An alignment of bat PRL sequences is shown in Fig. 1. Considerable sequence variation between species is apparent, with megabats being much less variable than microbats. However, the most striking aspect of the sequence comparison is the insertion of an additional sequence in vespertilionid bats. This occurs within exon 4, and clearly reflects an insertion of additional coding sequence into this exon rather than introduction of an additional intron, because the characteristic sequence motif at the start and end of an intron (GT....AG), required for processing, is not present. Furthermore, in several species analysis of transcriptomic databases showed that the inserted sequence is found in mRNA (see section 3.6). The length of the inserted sequence varies from 18 to 60 amino acids (aa), with the longer sequences (>18 aa) found in *Myotis* spp. Apart from this substantial insert in exon 4, the mature (excluding signal peptide) bat PRL sequences contain no additional insertions or deletions (indels). A substantial insertion of the kind seen in vespertilionid bats does not appear to have been reported for

PRL from any other mammalian or nonmammalian species (Ocampo Daza and Larhammar, 2018; Wallis, 2000). The nature of the inserted sequence is considered in detail below. It should be noted that for four of the vespertilionid species (*Myotis brandtii, M. davidii, M. lucifugus* and *Eptesicus fuscus*) predicted PRL CDS sequences are given in GenBank entries (respectively XM_005869979, XM_006757283, XM_006093744 and XM_008146068) in which the inserted sequence is interpreted as an intron and excluded. Reasons for thinking that this is an incorrect interpretation are given above. Revised prolactin CDS sequences for these four species have been deposited in the Third Party Annotation Section of the DDBJ/ENA/GenBank databases under the accession numbers TPA: BK010631-BK010634.

3.2. Variable rate of evolution of prolactin in bats

The bat PRL sequence alignment suggested variable rates of evolution (Fig. 1). This was examined further using the codeml programme and the corresponding CDS alignment for mature PRL, excluding the inserted sequence (Supplementary Fig.1). Analysis allowing assignment of dN/dS values independently to each branch indicated a markedly variable evolutionary rate (Fig. 2), with very variable branch lengths and dN/dS values. Application of the likelihood ratio test confirmed significant variation between branches and clades (Fig. 2). The overall rate of PRL evolution in megabats is very low (dN/dS = 0.062), similar to that for the outgroups (dog and rabbit, selected as having shown low evolutionary rate previously, dN/dS = 0.049). In most microbat groups the rate of PRL evolution (dN/dS) is much higher, but very variable. Thus it falls back to the slow rate seen for megabats in *Miniopterus natalensis* (dN/dS = 0.039) and in some cases markedly different dN/dS ratios were seen for closely related species pairs (e.g. *Desmodus rotundus,* dN/dS = 0.204/*Carollia brevicauda*, dN/dS = 0.843, Fig. 2). For some branches dN/dS > 1.0, but not significantly so, given the need for multiple test correction. Therefore, the accelerated evolution could not be shown formally to be due to positive selection rather than a loss of biological function (relaxed selection), though the latter seems unlikely on biological grounds.

A more discriminating test for positive selection is the branch-site method in codeml, which tests for variation in evolutionary rates between specific sites on designated branches (Zhang et al., 2005). Applying this method to the bat PRL alignment, positive selection was established for the

vespertilionid clade (dN/dS = 1.90; P<0.01; 75 sites identified as showing positive selection), and also for the *Myotis* clade (dN/dS = 3.01; P<0.05; 41 sites identified as showing positive selection) (Supplementary Table 2). Thus the accelerated evolution in these groups at least was a consequence of positive selection rather than relaxed selection.

3.3. The inserted sequence in vespertilionid bats

As indicated, the most striking novel feature observed in this study of bat PRL was the introduction of an additional sequence into the 4th exon in vespertilionid bats (Fig.1). The location of this insertion corresponds to the short loop between helices 3 and 4 in the 3D structure (see below). The insertion varies in length, from 18 aas in *Eptesicus*, *Pipistrellus* and *Lasiurus*, to 36-60 aas in the 4 *Myotis* species studied. It appears to have arisen as a series of internal duplications of an 18 nucleotide (nt) (6-aa) sequence, the original 6 aa sequence corresponding almost exactly to the short loop between helices 3 and 4. This repeating 6-aa unit is seen in each of the inserted regions (Fig. 3), though partially masked by the large number of substitutions that have been introduced subsequently. The inserted sequence could not be included in the codeml analysis because it varies in length between species. However, comparison of sequences of equivalent length shows that it is more variable than the adjacent sequence, suggesting that it too may be evolving under positive selection rather than relaxed selection.

The long insertion seen in vespertilionid bat PRL has no equivalent in any other species, and no apparent similarity could be detected with any other protein sequence. Its introduction appears to reflect a mechanism by which a novel repeat protein sequence can be built up rapidly during the course of evolution. Its function is not clear, but the observation that it and the rest of the PRL sequence in Vespertilionidae appears to be subject to positive selection suggests that it does have a specific, possibly novel, role. The short loop between helices 3 and 4 is not close to either receptor-binding site on PRL, and extending the length of this loop may not necessarily affect receptor binding (see below).

3.4. Bat prolactin receptor

The bat PRLR was investigated in order to assess to what extent the evolutionary changes seen in the hormone are reflected in changes in the receptor. Coevolution of PRL and its receptor has been reported in some other mammalian groups (Li et al., 2005)

Complete PRLR sequences were derived from genomic and transcriptomic databases for 19 bat species (Supplementary Table 1). For not all species were both PRL and PRLR sequences obtained, due to incomplete data. The PRLR sequences are aligned in Fig. 4. In many cases modest allelic polymorphism was observed, but in none was there evidence for duplication of the PRLR gene. As has been reported for other mammalian PRLR genes (Edery et al., 1989; Hu et al., 1999), the coding sequence was divided into 8 exons separated by 7 introns. The protein sequence is divided into 4 distinct domains, signal peptide, extracellular domain (ecd), single transmembrane domain (tmd) and intracellular domain (icd) (Fig. 4). Since these domains have quite distinct functions, and little structural interaction, they were considered separately for the evolutionary analysis with emphasis on the ecd and icd since signal peptide and tmd are quite short.

The ecd alignment (Fig. 4) shows that there is considerable variation between ecd sequences, with megabat sequences being less divergent than those of microbats, though this distinction is less marked than for PRL. The phylogenetic trees of Fig. 5 confirm this. Analysis using codeml (of the ecd CDS alignment given in Supplementary Fig. 2) gave an overall dN/dS of 0.371, and showed significant rate variation along branches ($P<0.01$, LRT), though this is less marked than seen for PRL. The rate of evolution for ecd in megabats is lower than that in microbats (Fig. 5), but in contrast to PRL the difference is not significant. Only for Vespertilionidae was the rate of evolution (dN/dS) significantly elevated (Fig. 5), but dN/dS did not exceed 1.0, so positive selection rather than relaxed selection could not be established.

The icd alignment (Fig. 4) shows substantial variation between species, but the pattern of this differs markedly from that seen for the ecd. Analysis using codeml (of the icd CDS alignment given in Supplementary Fig. 3) showed a high overall dN/dS (0.478), but no significant rate variation between lineages ($P > 0.1$; LRT). Scrutiny of the alignment suggested the presence of regions within the icd that were strongly conserved, and others that were very variable. The most notable conserved region is that immediately following the tmd, which contains box 1 and box 2, conserved motifs found in many cytokine receptors and required for biological function (DaSilva et al., 1994). Box 1 is crucial for association with protein kinase JAK2. Also fully conserved are 4 tyrosine residues in the C-terminal

half of the icd, likely to be subject to phosphorylation providing docking sites for other signalling molecules. These are the only Tyr residues completely conserved in the icd; short sequences around them are also strongly conserved. Haxholm et al. (2015) have proposed that the icd is largely unstructured, but includes 5 regions forming temporary helices (shown in Fig. 4). The first 2 of these (THD1-THD2) are included in the region close to the tmd, and are therefore strongly conserved. THD3 partially overlaps this region and is less conserved, and THD4 and THD5 are much less conserved. Haxholm et al. (2015) have also obtained evidence for three regions of the icd that associate with the plasma membrane (LD1-LD3); LD1 is included in the region close to the TMD, and is strongly conserved; LD2 and LD3 are very variable (Fig. 4). The pattern of variation seen in the icd accords with the view (Haxholm et al., 2015; Bugge et al., 2016) that the icd shows an open, relatively unstructured conformation, acting as a scaffold for binding of signalling molecules such as JAK2 kinase and STAT5. The conserved regions would then be the conserved specific binding sites for such molecules, while the regions between would act as spacers, with little requirement for specific, conserved sequence. It is also notable in this context that frequent indels are seen in the icd sequence (13 in the alignment of Fig. 4), in contrast to the complete lack of such indels in the ecd

3.5. Structural models of the PRL:PRLR$_2$ complex

Several 3D structures are available for PRL bound to the ecd of its receptor, including at least 2 for the putatively biologically active heterotrimer, PRL:PRLR$_2$ (Broutin et al., 2010; van Agthoven et al., 2010). The structure of the complex between human PRL and the rat PRLR ecd (van Agthoven et al., 2010) was used as the template for molecular modelling using SWISS-MODEL, to give homology models for several bat PRL:PRLR$_2$ complexes. Three of these are illustrated in Fig. 6. In all cases the model structures obtained matched the template closely, and Ramachandran plots were of a quality similar to that of the template.

Of particular interest is the location of the insert sequence seen in vespertlionid bat PRL. For PRL of the megabat *Pteropus alecto*, for which there is no inserted sequence, the short loop between helices 2 and 3 is seen to be well removed from the receptor binding sites (Fig. 6 top). For *Eptesicus fuscus*, with an inserted sequence of 18 residues, (Fig. 6 middle) and *Myotis brandtii*, with an insert of 60 residues, (Fig. 6 bottom) the inserts in the short loop are again seen to be far removed from the receptor

binding sites, and unlikely to interfere with receptor binding. It should be noted, however, that structures of the inserted sequences are unlikely to be very accurate, given that there is no equivalent to them in the template. The main conclusion from this structural study is that the insertions occur in a region compatible with retention of receptor binding.

The structural models also allow assessment of substitutions occurring during the course of evolution in relation to their positioning within the structure. The distribution of substitutions appeared to be non-random. Thus for the *Myotis* clade, the branch-site analysis (see above) identified 41 residues in PRL that appear to be subject to positive selection; these were distributed mainly in two regions, at the end of PRL distal from the receptor-binding sites and close to the inserted sequence, and in binding site 2 (Fig.7A and B). In the other vespertilionids, however, concentration of substitutions in these regions was not seen, and a large proportion of substitutions was seen at the end of PRL closest to the binding sites (Fig.7C and D). Such non-random distribution of substitutions accords with the view that PRL in vespertilionids is evolving under positive selection, and suggests that the specific adaptations involved vary among groups. In the PRLR ecd the distribution of substitutions occurred widely across the structure, with no concentration in specific regions.

3.6 Transcription of PRL and PRLR

Transcriptomic data is available for a number of bat species (summarized in Supplementary Table 1). No transcriptomes for pituitary gland (the main site for PRL synthesis) or placenta (the site for synthesis of PRL-like hormones in some mammalian groups) were specifically available. However, transcriptomes for brain from a number of species were available; in some of these PRL transcription was quite high, in others it was low, which may reflect varying inclusion of anterior pituitary gland in the brain material used. High levels of PRL transcription were also seen in trigeminal nuclei (in *Desmodus rotundus* and *Carollia brevicauda*), which has been shown to be a site of high PRL expression in rats (Diogenes et al., 2006), and in cochlea (*Rhinolophus macrotis, Hipposideros armiger*), a site of PRL expression in female mice (Marano et al., 2013). PRL transcription at low levels was seen in a number of peripheral tissues, as expected from studies on other mammals (Ben-Jonathan et al., 1996), but in none of these cases was a complete sequence derived. Transcriptomes confirmed the presence of the insertion sequence in PRL mRNA from several vespertilionid species,

and its absence in many other bat species. Evidence for alternative splicing of PRL mRNA precursor was seen in three species (*Aselliscus stoliczkanus, H. armiger, R.macrotis*), with inclusion or exclusion of one codon (GCA, encoding Ala) at the exon1/exon2 junction (Supplementary Table 1B).

Moderate to high expression of PRLR was seen in many bat tissues, from a considerable number of species, with highest abundance observed in liver, brain, cochlea, sub-mandibular gland and testis. Some evidence was seen for alternative splicing at the first splice junction between coding exons, with a variant with 5 nt excluded at this site, and another including 60 nt of the intron. The significance of these variants is not clear, but they were much less abundant that the main form, and any resulting variation at the protein level would probably only involve the signal peptide. Alternative splicing involving additional upstream non-coding exons was not investigated.

3.7 Conclusions

PRL evolution in Chiroptera shows episodic evolution, but the nature of this is rather different from that seen in several other mammalian groups, including primates and cetartiodactyls, where a single episode of rapid change contrasts with an otherwise very slow basal evolution (Wallis, 2000; Wallis et al., 2005). In Chiroptera very slow evolution is seen in megabats, but rapid evolution appears to be sustained throughout microbats, with a few exceptions such as the lineage leading to *Miniopterus*. The rapid evolution is particularly marked in the Vespertilionidae, the largest bat family, with a world-wide distribution, where it accompanies the insertion of an additional peptide sequence of 18-60 amino acids in the short loop between helices 2 and 3. On some branches within Vespertilionidae $dN/dS > 1$, suggesting selective, adaptive evolution. Branch-site analysis confirmed adaptive evolution within Vespertilionidae and identified specific residues likely to be subject to positive selection. Notably, these sites show an apparently non-random distribution in the 3D structure. The very rapid evolution of PRL in Vespertilionidae is emphasized by comparison with the sequence of a reptilian PRL (alligator; Noso et al., 1992; Wallis 2000). The sequence of a megabat PRL (*Pteropus alecto*) is more similar to PRL of the alligator (*Alligator mississippiensis*) than to that of a vespertilionid bat (*Myotis lucifugus*) (Fig. 1), although lineages leading to microbats and megabats diverged 60-65 million years ago (MYA) (Amador et al., 2018) while those leading to reptiles and mammals diverged about 310 MYA (Kumar and Hedges, 1998).

The inserted sequence in PRL in vespertilionids also appears to be undergoing rapid adaptive evolution though quantitative evaluation is difficult because of the variable length of the inserted sequences. The insert sequence shows a repetitive structure, reflecting internal duplication of the 6 amino acids of the short loop. Such an insert is not seen in PRL from other mammalian groups, or indeed other vertebrates, and appears to illustrate the origin and evolution of a completely novel sequence. Elucidation of the function associated with the insert will require physiological and biological studies; its location far-removed from the receptor-binding sites suggests that it may not affect receptor-binding.

The episodic evolution seen in bat PRL is reflected, though to a lesser extent, in the PRLR ecd, suggesting coevolution of PRL and its receptor as has been detected in some other groups (Li et al., 2005). The evolution of the icd shows a completely different pattern of evolution, with no evidence for variable rates of evolution on differing bat lineages, but marked variation in sequence conservation along the sequence, with strongly conserved regions alternating with very variable ones. It seems likely that this reflects the rather open structure of the ecd, providing strongly conserved binding sites for intracellular signalling molecules, joined by very variable linker regions.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Agnarsson, I., Zambrana-Torrelio, C.M., Flores-Saldana, N.P., May-Collado, L.J., 2011. A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia). PLOS Currents Tree of Life. Feb 4 . Edition 1. doi: 10.1371/currents.RRN1212.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Amador, L.I., Moyers Arévalo, R.L., Almeida, F.C., Catalano, S.A., Giannini, N.P., 2018. Bat sytematics in the light of unconstrained analyses of a comprehensive molecular supermatrix. J.

Mammal. Evol. 25, 37-70.

Ben-Jonathan, N., Mershon, J.L., Allen, D.L., Steinmetz, R.W., 1996. Extrapituitary prolactin: distribution, regulation, functions, and clinical aspects. Endocr. Rev. 17, 639-669.

Broutin, I., Jomain, J.-B., Tallet, E., van Agthoven, J., Raynal, B., Hoos, S., Kragelund, B.B., Kelly, P.A., Ducruix, A., England, P., Goffin, V., 2010. Crystal structure of an affinity-matured prolactin complexed to its dimerized receptor reveals the topology of hormone binding site 2. J. Biol. Chem. 285, 8422–8433.

Bugge, K., Papaleo, E., Haxholm. G.W., Hopper, J.T.S., Robinson, C.V., Olsen, J.G., Lindorff-Larsen, K., Kragelund, B.B., 2016. A combined computational and structural model of the full-length human prolactin receptor. Nature Comm. 7, art. 11578.

Carter, A.M., 2012. Evolution of placental function in mammals: the molecular basis of gas and nutrient transfer, hormone secretion, and immune responses. Physiol. Rev. 92, 1543-1576.

Clarke, W. C., Bern, H.A., 1980. Comparative endocrinology of prolactin, in: Li, C.H. (Ed.), Hormonal Proteins and Peptides, Academic Press, New York, Vol. 8, pp. 105-197.

Clevenger, C.V., Gadd, S.L., Zheng, J., 2009. New mechanisms for PRLr action in breast cancer. Trends Endocrinol. Metab. 20, 223-229.

Dammann, P., 2017. Slow aging in mammals - lessons from African mole-rats and bats. Seminars Cell Dev. Biol. 70, 154-163.

DaSilva, L., Howard, O.M.Z., Rui, H., Kirken, R.A., Farrar, W.L., 1994. Growth signaling and JAK2 association mediated by membrane-proximal cytoplasmic regions of prolactin receptors. J. Biol. Chem. 269, 18267-18270.

Diogenes, A., Patwardhan, A.M., Jeske, N.A., Ruparel, N.B., Goffin, G., Akopian, A.N., Hargreaves, K.M., 2006. Prolactin modulates TRPV1 in female trigeminal sensory neurons. J. Neuroscience 26, 8126-8136.

Edery, M., Jolicoeur, C., Levi-Meyrueis, C., Dusanter-Fourt, I., Pétridou, B., Boutin, J-M., Lesueur, L., Kelly, P.A., Djiane, J., 1989. Identification and sequence analysis of a second form of prolactin receptor by molecular cloning of complementary DNA from rabbit mammary gland. Proc. Nat. Acad. Sci. USA. 86, 2112-2116.

Forsyth, I.A., Wallis, M., 2002. Growth hormone and prolactin - molecular and functional evolution. J. Mammary Gland Biol. Neoplasia 7, 291-312.

Haxholm, G.W., Nikolajsen, L.F, Olsen, J.G., Fredsted, J., Larsen, F.H., Goffin, V., Pedersen, S.F., Brooks, A.J., Waters, M.J., Kragelund, B.B., 2015. Intrinsically disordered cytoplasmic domains of two cytokine receptors mediate conserved interactions with membranes. Biochem. J. 468, 495-506.

Horseman, N.D., Gregerson, K.A., 2014. Prolactin actions. J. Mol. Endocrinol. 52, R95-R106.

Hu, Z.-Z., Zhuang, L., Meng, J., Leondires, M., Dufau, M.L., 1999. The human prolactin receptor gene structure amd alternative promoter utilization: the generic promoter hPIII and a novel human promoter hP_N. J. Clin. Endocrinol. Metab, 84, 1153-1156.

Kumar, S., Hedges, S.B., 1998. A molecular timescale for vertebrate evolution. Nature, 392, 917-920.

Li, Y., Wallis, M., Zhang, Y-P., 2005. Episodic evolution of prolactin receptor gene in mammals: coevolution with its ligand. J. Mol. Endocrinol. **35**, 411-419.

Maddison, W.P., Maddison, D.R., 2018. Mesquite: A Modular System for Evolutionary Analysis. Version 3.51. Available from: http://www.mesquiteproject.org.

Marano, R.J., Tickner, J., Redmond, S.L., 2013. Prolactin expression in the cochlea of aged BALB/c mice is gender biased and correlates to loss of bone mineral density and hearing loss. PLoS ONE 8, e63952.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418-426.

Noso, T., Swanson, P., Lance, V.A., Kawauchi, H., 1992. Isolation and characterization of glycosylated and non-glycosylated prolactins from alligator and crocodile. Int. J. Pept. Prot. Res. 39, 250–257.

Ocampo Daza, D., Larhammar, D., 2018. Evolution of the growth hormone, prolactin, prolactin 2 and somatolactin family. Gen. Comp. Endocrinol. 264, 94-112.

Shillingford, J.M., Hennighausen, L., 2001. Experimental mouse genetics - answering fundamental questions about mammary gland biology. Trends Endocrinol. Metab. 12, 402-408.

Teilum, K., Hoch, J.C., Goffin, V., Kinet, S., Martial, J.A., Kragelund, B.B., 2005. Solution structure of human prolactin. J. Mol. Biol. 351, 810-823.

Truong, A.T., Duez, C., Belayew, A., Renard, A., Pictet, R., Bell, G.I., Martial, J.A., 1984. Isolation and characterization of the human prolactin gene. EMBO J. 3. 429–437.

van Agthoven, J., Zhang, C., Tallet, E., Raynal, B., Hoos, S., Baron, B., England, P., Goffin, V., Broutin, I., 2010. Structural characterization of the stem-stem dimerization interface between

prolactin receptor chains complexed with the natural hormone. J. Mol. Biol. 404, 112–126.

Wallis, M., 1981. Molecular evolution of pituitary growth hormone, prolactin and placental lactogen: a protein family showing very variable rates of evolution. J. Mol. Evol. 17, 10-18.

Wallis, M., 1994. Variable evolutionary rates in the molecular evolution of mammalian growth hormones. J. Mol. Evol. 38, 619-627.

Wallis, M., 2000. Episodic evolution of protein hormones: molecular evolution of pituitary prolactin. J. Mol. Evol. 50, 465-473.

Wallis, M., 2008. Mammalian genome projects reveal new growth hormone (GH) sequences. Characterization of the GH-encoding genes of armadillo (*Dasypus novemcinctus*), hedgehog (*Erinaceus europaeus*), bat (*Myotis lucifugus*), hyrax (*Procavia capensis*), shrew (*Sorex araneus*), ground squirrel (*Spermophilus tridecemlineatus*), elephant (*Loxodonta africana*), cat (*Felis catus*) and opossum (*Monodelphis domestica*). Gen. Comp. Endocrinol. 155, 271-279.

Wallis, M., 2009. Prolactin in the Afrotheria: characterization of genes encoding prolactin in elephant (*Loxodonta africana*), hyrax (*Procavia capensis*) and tenrec (*Echinops telfairi*). J. Endocrinol. 200, 233-240.

Wallis, O.C., Mac-Kwashie, A.O., Makri, G., Wallis, M., 2005. Molecular evolution of prolactin in primates. J. Mol. Evol. **60**, 606-614.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296-W303.

Yang, Z., 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. 15, 568-573.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evo. 24. 1586–1591.

Yang, Z., Nielsen, R., 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J. Mol. Evol. 46, 409-418.

Yang, Z., Wong, W.S.W., Nielsen, R., 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22, 1107-1118.

Zhang, J., Nielsen, R., Yang, Z., 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22, 2472-2479.

Figure Legends

Fig. 1. Alignment of bat prolactin sequences. Sequences of rabbit, human, dog and alligator PRLs are included as outgroups. The sequence of rabbit PRL is shown in full on the top line. For sequences below this, . indicates a residue identical to rabbit, - indicates a gap. The signal peptide is shaded grey and the position of the inserted sequence found in vespertilionid bats is indicated. Numbers at the right indicate differences from rabbit for: (a) signal peptide, (b) mature protein, (c) total (excluding inserted sequences). Location of exons in the corresponding nucleotide sequence is indicated on the bottom line. Abbreviations for species names: Ocu, *Oryctolagus cuniculus*; Hsa, *Homo sapiens*; Cfa, *Canis familiaris*; Pal, *Pteropus alecto*; Pva, *Pteropus vampyrus*; Ast, *Aselliscus stoliczkanus*; Ehe, *Eidolon helvum*; Rae, *Rosettus aegypticus*; Mly, *Megaderma lyra*; Hsi, *Hipposideros armiger*; Rfe, *Rhinolophus ferrumequinum*; Rsi, *Rhinolophus sinicus*; Rma, *Rhinolophus macrotis*; Mna, *Miniopterus natalensis*; Ppa, *Pteronotus parnellii*; Dro, *Desmodus rotundus*; Cbr, *Corollia brevicauda*; Efu, *Eptesicus fuscus*; Ppy, *Pipistrellus pygmaeus*; Lbo, *Lasiurus borealis*; Mlu, *Myotis lucifugus*; Mbr, *Myotis brandtii*; Mru, *Myotis rufonige*; Mda, *Myotis davidii*; Ami, *Alligator mississippiensis*.

Fig. 2. Phylogenetic trees for bat prolactin. Based on analysis of an alignment of CDS sequences for mature PRL (excluding signal peptides and inserted sequences in Vespertilionidae) using codeml and a defined tree. Trees based on synonymous sites (dS, top) and nonsynonymous sites (dN, bottom) are shown. Numbers at the right-hand side of the dN tree are dN/dS values for the clades indicated. Difference from the mean dN/dS value for the tree (0.356) are shown as * $P<0.05$, *** $P<0.001$ (likelihood ratio test). Scale bars indicate nucleotide substitutions/codon. Abbreviations for species names are as for Fig. 1.

Fig. 3. Hexapeptide repeats within the inserted sequences in vespertilionid bat prolactin. For each species, the top sequence is the hexapeptide preceding the inserted sequence. Sequences below this represent the insert broken into hexapeptides. For full sequences of the inserts see Fig. 1.

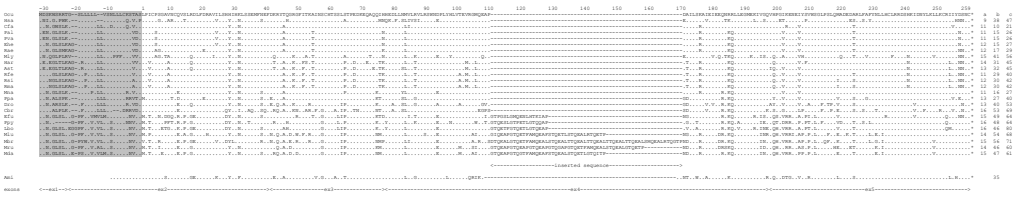
Fig. 4. Alignment of bat prolactin receptor sequences. Sequences of rabbit, human and dog PRLRs are

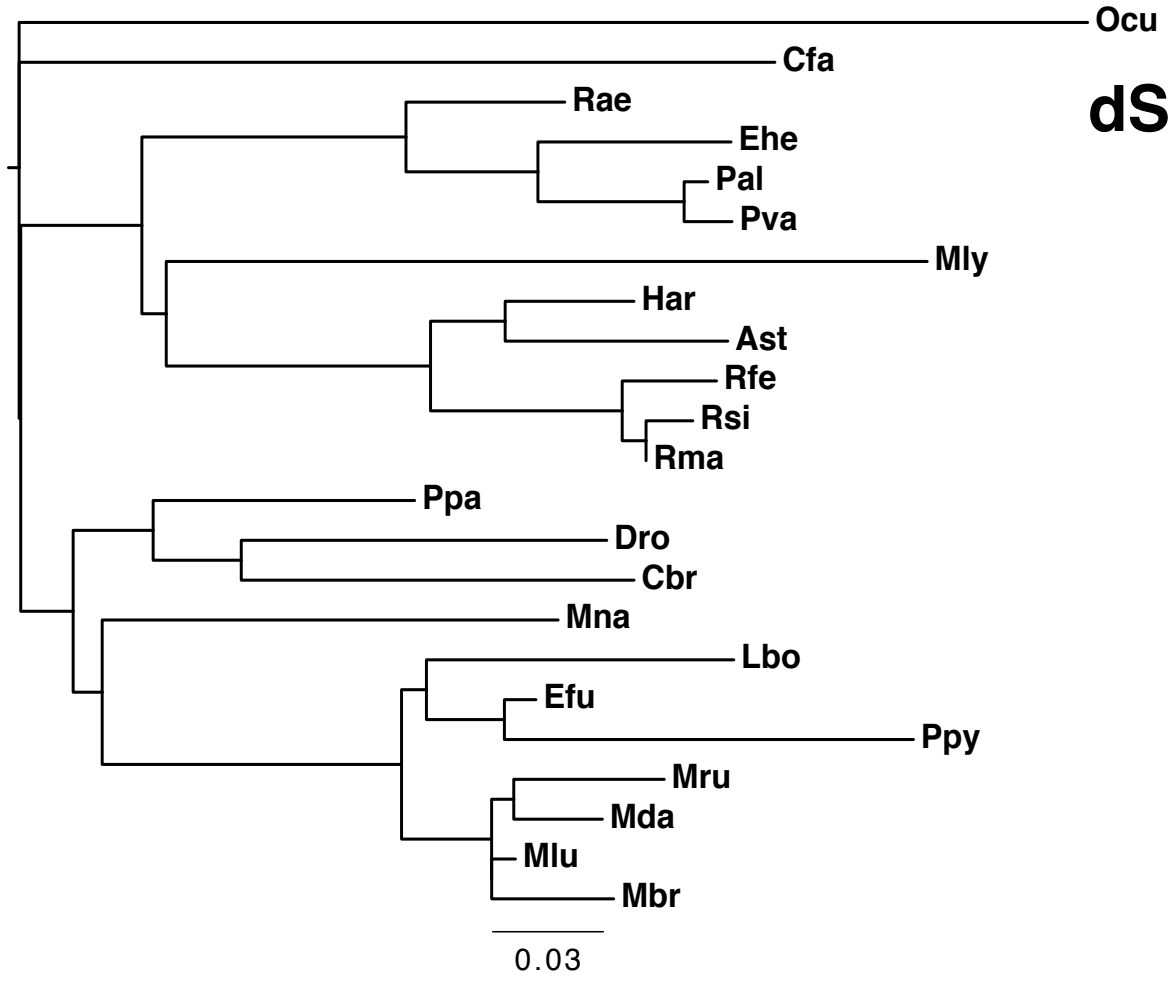
included as outgroups. The sequence of rabbit PRLR is shown in full on the top line. For sequences below this, . indicates a residue identical to rabbit, - indicates a gap. The signal peptide and transmembrane domain are shaded grey. Potential functional components of the icd are indicated: box 1 and box 2, TDH1-5, LD1-3. Numbers at the right indicate differences from rabbit (including indels) for: (a) signal peptide, (b) ecd, (c) tmd, (d) icd (e) total. Abbreviations for species names are as for Fig. 1 plus Mle, *Murina leucogaster*.

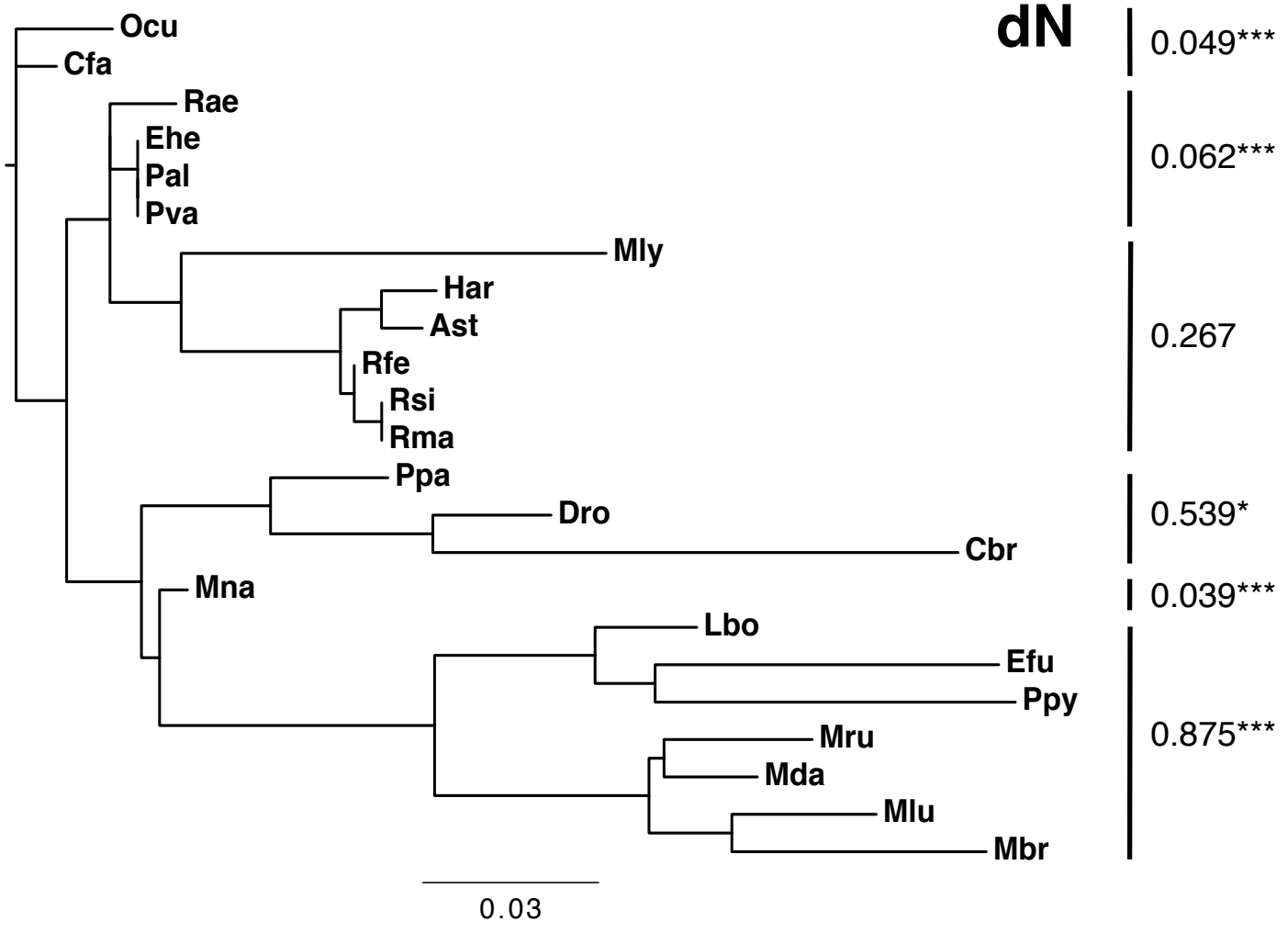
Fig. 5. Phylogenetic trees for bat prolactin receptor ecd. Based on analysis of an alignment of CDS sequences using codeml and a defined tree. Trees based on synonymous sites (dS, top) and nonsynonymous sites (dN, bottom) are shown. Numbers at the right-hand side of the dN tree are dN/dS values for the clades indicated. Difference from the mean dN/dS value for the tree (0.371) are shown as ** $P<0.01$, *** $P<0.001$ (likelihood ratio test). Scale bars indicate nucleotide substitutions/codon. Abbreviations for species names are as for Figs. 1 and 4.

Fig. 6. Molecular models for PRL:$PRLR_2$ from 3 bat species. In each case, the model was constructed using the SWISS-MODEL homology modelling server, with the human PRL:rat $RPRL_2$ structure (3npz) as template. The view is looking down towards the membrane. PRL is in green except for the loop between helices 2 and 3 plus inserted sequence which is in blue, PRLR ecds are in cyan and magenta; other receptor domains are not included.

Fig. 7. Distribution of substitutions within the homology models of bat prolactin. The substitutions identified as potentially involving positive selection were mapped onto the PRL:$PRLR_2$ homology models for *Myotis lucifugus* (A and B; substitutions in *Myotis* clade) or *Eptesicus fuscus* (C and D; substitutions in Vespertilionidae excluding *Myotis*). A and C are viewed looking down, towards the membrane, B and D are viewed looking up, away from the membrane. The PRLR ecd chains are shown in cartoon format, chain B cyan, chain C magenta. PRL is shown in space filling format, blue, with the loop between helices 3 and 4 (including insert sequence) shown in dark blue. PRL residues identified as potentially positively selected in the branch-sites analysis (excluding inserted sequence) are shown as orange (residues in binding site 1), red (residues in binding site 2), purple (in binding sites 1 and 2) or yellow (other residues).




dS
Ocu
Cfa
Rae
Ehe
Pal
Pva
Mly
Har
Ast
Rfe
Rsi
Rma
Ppa
Dro
Cbr
Mna
Lbo
Efu
Ppy
Mru
Mda
Mlu
Mbr
0.03


dN
Ocu
Cfa
Rae
Ehe
Pal
Pva
Mly
Har
Ast
Rfe
Rsi
Rma
Ppa
Dro
Cbr
Mna
Lbo
Efu
Ppy
Mru
Mda
Mlu
Mbr
0.03
0.049***
0.062***
0.267
0.539*
0.039***
0.875***

Eptesicus fuscus

GMQEIP
GTPGSL
GMQENL
HTKIAP

Pipistrellus pygmaeus

GWQETP
GTQESL
GTPETL
GTQQAP

Lasiurus borealis

GMQEAP
GTQETF
GTQETL
GTQEAP

Myotis lucifugus

AIQEAP
GIQEAP
GTQETF
AMQEAF
STQETL
STQEAL
RTQETP

Myotis brandtii

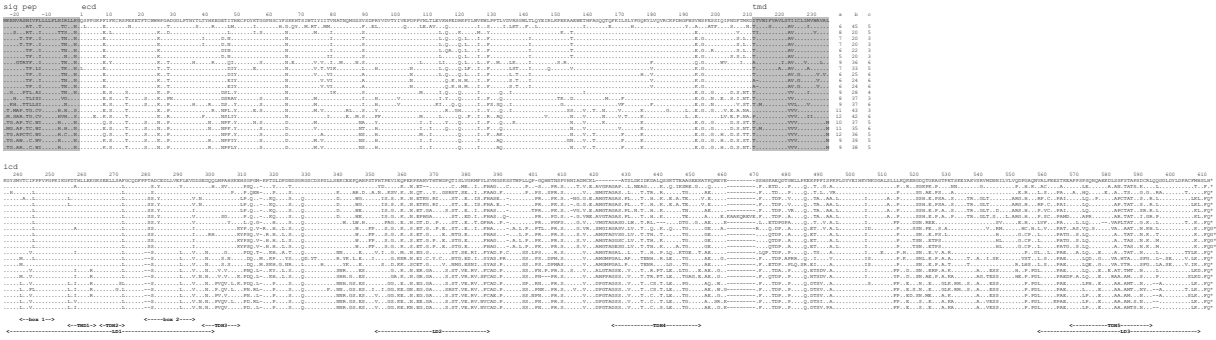
AMREAS
DTQKAL
GTQETF
AMQEAL
STQEAL
TTQEAL
TTQEAL
TTQEAL
TTQEAL
SMQEAL
RTQGTP

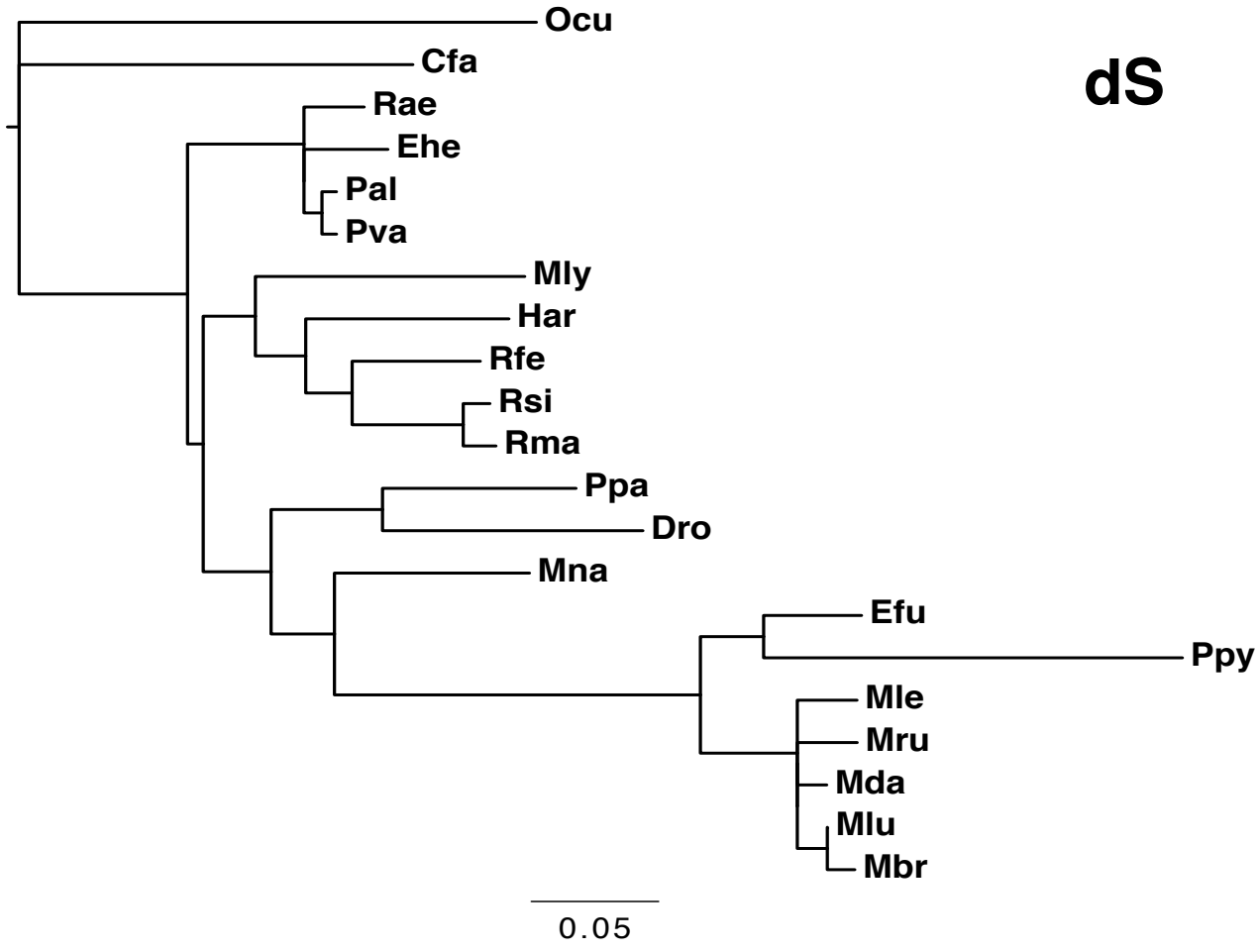
Myotis rufoniger

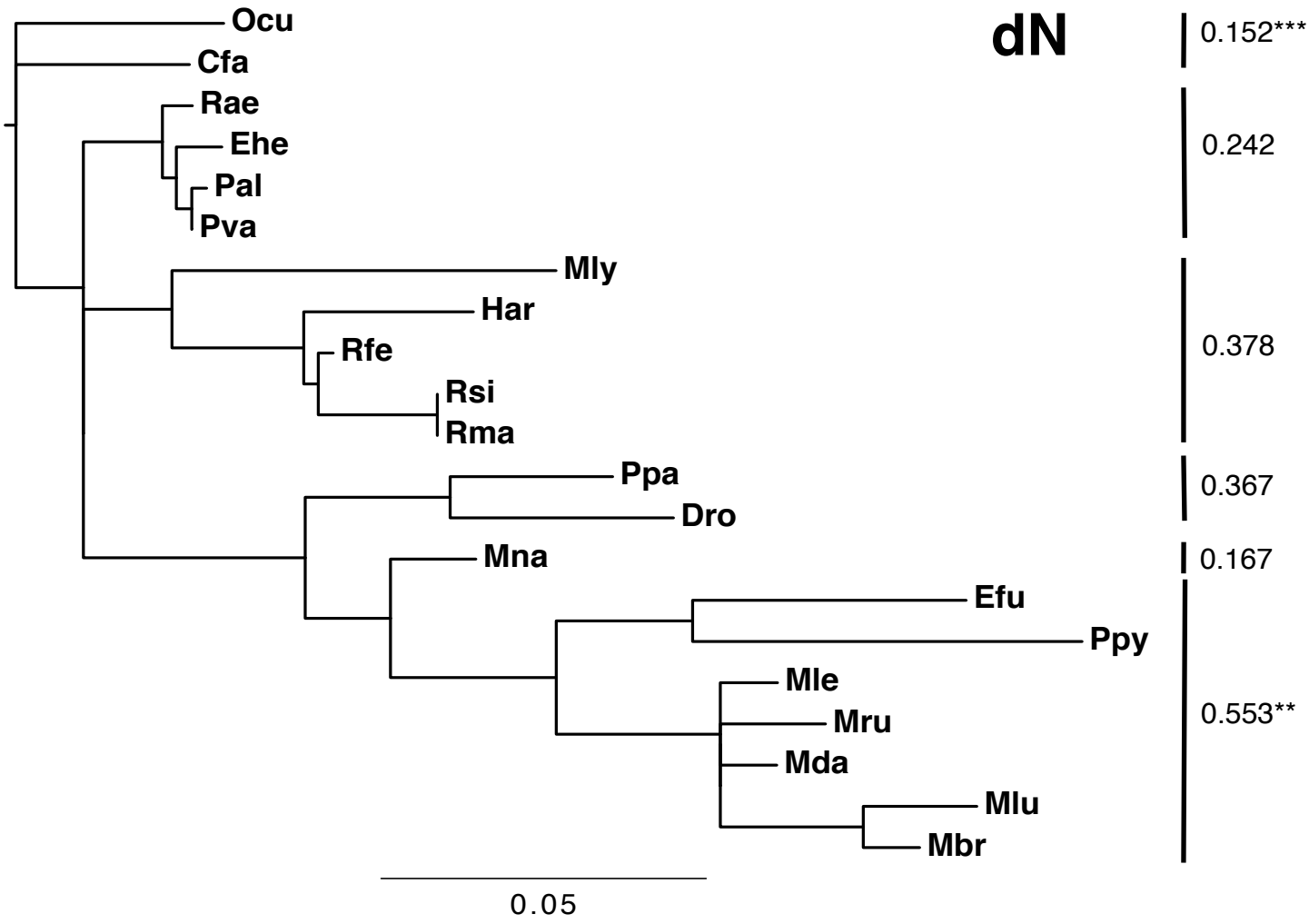
SIQEAP
GTQEAP
GTQEAP
GTQEAP
GTQGAP
GTQETF
AMQEAL
STQEAL
GTQETP

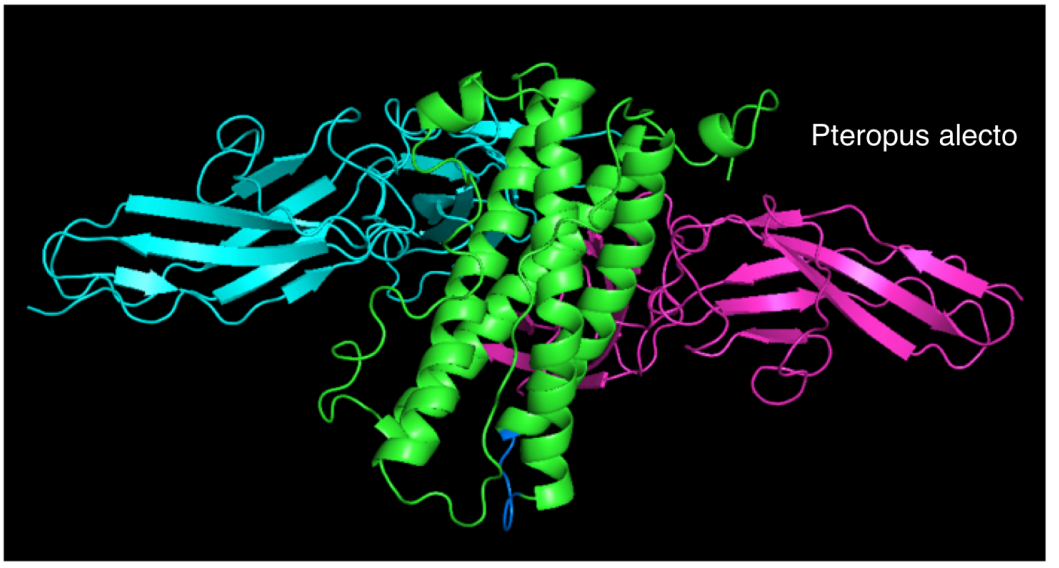
Myotis davidii

AIQEAP
GTQEAP
GTQGTF
AMQEAF
STQEAL
STQETL
GTQITP


sig pep
ecd
tmd
icd
TM1
TM2
TM3
TM4
TM5
TM6
TM7
LE1
LE2
LE3

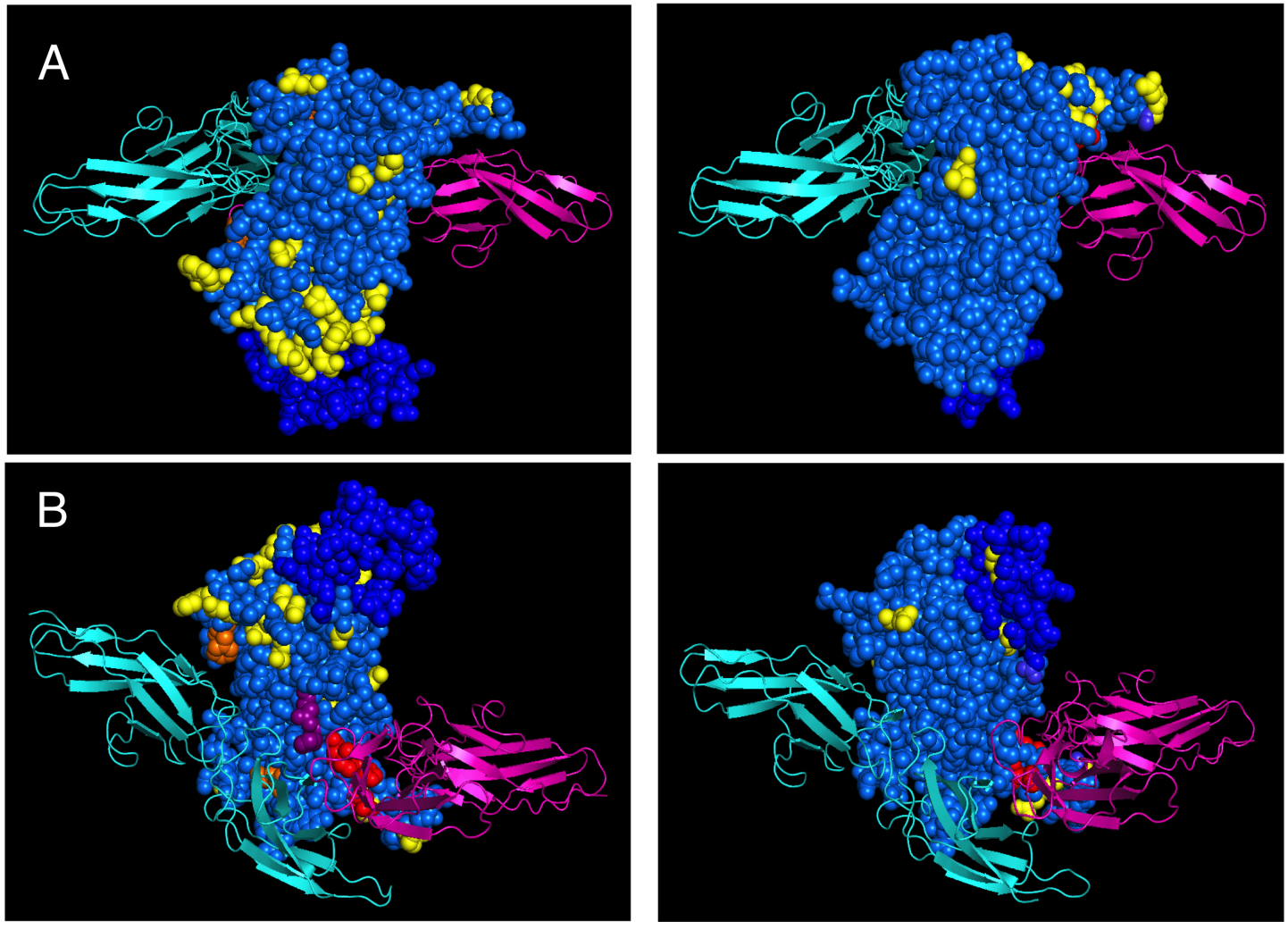

dS
Ocu
Cfa
Rae
Ehe
Pal
Pva
Mly
Har
Rfe
Rsi
Rma
Ppa
Dro
Mna
Efu
Ppy
Mle
Mru
Mda
Mlu
Mbr
0.05


dN
Ocu
Cfa
Rae
Ehe
Pal
Pva
Mly
Har
Rfe
Rsi
Rma
Ppa
Dro
Mna
Efu
Ppy
Mle
Mru
Mda
Mlu
Mbr
0.05
0.152***
0.242
0.378
0.367
0.167
0.553**


Pteropus alecto


Eptesicus fuscus


Myotis brandtii


A
B

A) Branch-site analysis for *Myotis* clade

lnL = -3954.894481

dN/dS (w) for site classes (K=4)

site class	0	1	2a	2b
proportion	0.50832	0.18266	0.22734	0.08169
background w	0.16319	1.00000	0.16319	1.00000
foreground w	0.16319	1.00000	3.01081	3.01081

Bayes Empirical Bayes (BEB) analysis
Positive sites for foreground lineages Prob(w>1):

3	I	0.648
6	S	0.715
7	G	0.712
14	S	0.947
17	D	0.640
21	R	0.794
24	I	0.700
27	H	0.659
29	I	0.736
31	K	0.612
38	N	0.859
41	D	0.630
43	R	0.668
47	G	0.975*
49	G	0.948
51	I	0.732
53	K	0.744
74	Q	0.830
84	V	0.564
89	R	0.746
100	T	0.564
104	G	0.980*
105	M	0.733
106	Q	0.722
109	P	0.707
111	A	0.557
121	E	0.919
123	N	0.672
136	Q	0.580
144	N	0.764
147	Y	0.886
154	P	0.810
155	S	0.696
157	Q	0.797
158	M	0.690
161	E	0.780
162	D	0.684
168	F	0.754
170	N	0.999**
172	L	0.725
188	L	0.959*

<u>For Null (w constrained to 1.0)</u>

lnL = -3959.046770

dN/dS (w) for site classes (K=4)

site class	0	1	2a	2b
proportion	0.35595	0.13292	0.37216	0.13897
background w	0.15773	1.00000	0.15773	1.00000
foreground w	0.15773	1.00000	1.00000	1.00000

LRT test:

2xΔlnL = 8.32; 2 degrees of freedom; P<0.05

tree for this analysis:

```
(Ocu,Cfa,(((Rae,(Ehe,(Pal,Pva))),(Mly,((Har,Ast),(Rfe,(Rsi,Rma)))
)),((Ppa,(Dro,Cbr)),(Mna,((Lbo,(Efu,Ppy)),((Mru,Mda),(Mlu,Mbr))$1
)))));
```

B) Branch-site analysis for Vespertilionidae clade

lnL = -3939.697653

dN/dS (w) for site classes (K=4)

site class	0	1	2a	2b
proportion	0.54940	0.07893	0.32498	0.04669
background w	0.16195	1.00000	0.16195	1.00000
foreground w	0.16195	1.00000	1.89670	1.89670

Bayes Empirical Bayes (BEB) analysis
Positive sites for foreground lineages Prob(w>1):

1	L	0.796
3	I	0.911
6	S	0.816
7	G	0.949
8	A	0.957*
9	V	0.973*
10	N	0.982*
12	Q	0.760
14	S	0.989*
17	D	0.991**
21	R	0.693
27	H	0.890
29	I	0.510
34	S	0.645
38	N	0.972*
41	D	0.540
42	K	0.574
43	R	0.649
45	T	0.560
46	Q	0.945
47	G	0.910
49	G	0.889
51	I	0.631
53	K	0.602
54	A	0.544

```
    57 S 0.529
    62 S 0.808
    63 L 0.898
    74 Q 0.734
    75 I 0.584
    76 H 0.603
    77 H 0.909
    83 M 0.681
    84 V 0.875
    86 R 0.559
    89 R 0.556
    97 H 0.611
   100 T 0.535
   104 G 0.951*
   105 M 0.987*
   106 Q 0.657
   108 A 0.810
   109 P 0.660
   110 D 0.610
   111 A 0.892
   114 S 0.957*
   121 E 0.875
   127 L 0.982*
   132 K 0.946
   135 G 0.996**
   136 Q 0.900
   139 P 0.994**
   142 K 0.694
   143 E 0.961*
   144 N 0.593
   146 I 0.898
   147 Y 0.700
   148 S 0.672
   149 V 0.987*
   151 S 0.761
   153 L 0.653
   154 P 0.745
   155 S 0.670
   157 Q 0.670
   158 M 0.679
   161 E 0.635
   166 F 0.615
   168 F 0.952*
   170 N 0.993**
   172 L 0.551
   173 H 0.619
   182 I 0.623
   188 L 0.902
   192 R 0.742
   196 D 0.553
```

For Null (w constrained to 1.0)

lnL = -3945.680906

dN/dS (w) for site classes (K=4)

site class 0 1 2a 2b

proportion	0.47101	0.07267	0.39532	0.06100
background w	0.15095	1.00000	0.15095	1.00000
foreground w	0.15095	1.00000	1.00000	1.00000

LRT test:

2xΔlnL = 11.96; 2 degrees of freedom; P<0.01

tree for this analysis:

```
(Ocu,Cfa,(((Rae,(Ehe,(Pal,Pva))),(Mly,((Har,Ast),(Rfe,(Rsi,Rma)))
)),((Ppa,(Dro,Cbr)),(Mna,((Lbo,(Efu,Ppy)),((Mru,Mda),(Mlu,Mbr)))$
1))));
```